

Bioethical Need for Generative Artificial Intelligence (AI)

August 27, 2023

By Joseph Erban

In a recent statement by The Center for AI Safety, a group of leading thinkers warns us of the need of: “mitigating the risk for extinction from AI should be a global priority.” This was followed by a meeting between US president Joe Biden and leading representatives of corporations currently developing AI technologies to come up with protective measures concerning such evolving technologies. The outcome of the meeting effectively left industry to regulate itself.

Human intelligence and now AI have the capacity to do good deeds, such as saving lives, but can potentially cause existential harm, including potentially killing us. The issue facing humanity regarding AI cannot be starker: choosing technologies that promote life and health or those which can cause harm, including extinction. History can nevertheless inform us as to what might be needed in order to render such technologies less threatening or harmful.

In maintaining cooperation, social in-groups have developed shared principles confined to ethical or religious rules, or social norms meant to promote how in-group members ought to function in various situations. Such values and norms are based on groups’ members willingness to accept such ethical or moral precepts, behaviors that are often culturally transmitted horizontally to in-group living members, or vertically, when such behaviors are passed intergenerationally.

Given the globally divergent moral and ethical landscape currently available demands that individuals either consciously or by social learning acquire a moral system requiring them to navigate through their proper society.

Even within a particular in-group such as in Canada, let alone other diverse cultures, contradictory moral views are nevertheless upheld by various groups. Diverse people may, therefore, view what is right or wrong, what is harmful or what is beneficial in opposite polls. Killing X might be perceived as morally justifiable by one group, and unjust by others.

Moral reasoning and action are derived from one’s moral beliefs. Implementing an ethics for AI will not be different, since it is humans, often working for the tech sector, who will program such algorithms. Moreover, AI programmers harbor divergent moral principles due to peoples’ diverse moral beliefs, their cultural background and upbringing. For translational tech corporations the aim is to maximize profit, with only enforceable regulation, should they exist, setting their limits.

In order for AI to reproduce ethical human moral reasoning and action will require adopting ethical guidelines as frameworks that are acceptable, by designers, developers, users of such

systems and their communities.

Given the current ethical relativism we encounter world-wide today, AI designer, developers and society would have to choose acceptable guidelines, legal codes, and policies, and tailor them specifically with programmed rules that are acceptable to those developing AI technologies, their consumers and society at large.

Ethical concerns for AI potential to cause harm can learn a great deal from the biomedical communities that have been dealing with ethical dilemmas arising from novel technologies entering healthcare setting and their communities. After all, bioethics is precisely the ethical discipline that deals with the healthcare ethics of life and death. It is accepted as a framework for those working in the medical and health sciences, as well as by those researchers using humans as their subjects, such as in drug trials.

Therefore, one possible solution for an AI ethics might be found by adopting biomedical ethical guidelines, frameworks and principles as learning datasets that have been generally accepted by the biomedical communities and can be adapted for a broader ethics for AI as well, since both are meant to guide healthcare providers, thus reducing ethical relativism found in a world of divergent ethical belief systems. The result can be in an ethical harmonization for AI as well, yet adaptable to specific circumstances, including healthcare, which it must do, if it is to be used in such settings.

A recent report issued by the World Health Organization entitled *“Ethics and Governance of Artificial Intelligence for Health”* outlines ethical principles it states as in need of being incorporated within AI technologies for healthcare [1]. These could serve as a blueprint not only for an ethic for AI in healthcare, but to ensure human safety, also in research setting that use humans as their subjects, as in the case in clinical trials, or with other therapeutics, prior to them being released into the market place.

One principle mentioned in the WHO’s document is autonomy: humans control decision making at all levels in health. Human autonomy entails truth-telling of information accurately and comprehensibly for effective treatment coupled with informed consent; respecting patient privacy and confidentiality empowering decision-making and data protection.

The WHO document notes the need to “promote human well-being”, also known as the beneficence principle and the non-maleficent principle stipulating the avoidance of harm - mental or physical.

The transparency, explicability and intelligibility principles are meant to ensure that the sources of information used as datasets in the design of AI are documented and available for public scrutiny and is clearly explained and debated publicly, with community participation, in such AI programming design and its eventual deployment. Industry would probably contest this requirement based on protecting property or intellectual rights. Therefore, laws concerning AI and ownership of intellectual property would have to be amended to

accommodate such ethical considerations.

Another principle mentioned requires AI to foster responsibility and accountability to its users including the development, implementation and regulatory policies that ensures ongoing safety through human oversight, the so call ‘human warranty’ or supervision of AI technologies by *all* shareholders.

Industry current position is to set up a third party to check for flaws. Just as is the case for many technological or therapeutic innovations first developed and release into the market place, these can be used in numerous and unexpected ways that third parties may overlook, as is the case with released new medications or device that can result in adverse effects. Think e-cigarettes, for example.

Given human biases, AI technologies ought to be inclusive and equitably available thus avoiding discrimination and protecting human rights for all.

The WHO's document suggests that AI technologies be responsive and sustainable. This requires monitoring ongoing assessment of safety and efficacy of AI technology, as well as incorporating it within governments existing environmentally sustainable goals.

WHO's guideline is a call for an ethical generative artificial intelligence in healthcare. Still, let us hope that regulators have the wisdom to come to a global consensus regarding adopting an ethics for all generative AI algorithms that will prevent future harms, which will arise if we hastily release novel technologies into the market place. Otherwise we run the risk of abrogating our moral responsibilities to ensure safety and beneficence, while avoiding unnecessary harm caused by of the addictive pursue of profit and power that the tech industry is currently chasing.

[1] WHO: *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization; 2021

<file:///C:/Users/cug05467/Downloads/9789240029200-eng.pdf>. Accessed August 28,2023